State Library
OF WESTERN AUSTRALIA

Creating and keeping your

DIGITAL TREASURES

a user guide

Updated December 2010

Copyright: Creative Commons, Noncommercial share alike
http://creativecommons.org/licenses/by-nc-sa/2.5/au/

Creating and Keeping your Digital Treasures: A User Guide

This document provides an overview of standards for users who wish to create or convert digital objects. The standards described in this document are provided as a guide, and focus on long term access and preservation of objects. It is anticipated that where required, you will conduct your own research. A number of links have been included in **Appendix 2** as a starting point for your research.

The State Library of Western Australia is committed to expanding its collection and documenting Western Australian culture. In the modern (digital) environment, an increasing proportion of objects being collected are born digital, which means that they are digital at the point of creation. The prevalence of affordable recording technologies are increasing the ability for a large proportion of Western Australians to create, capture and store objects in digital formats. Reformatting hardware (scanners, etc) is also prevalent, allowing users to easily reformat non-digital objects (e.g. convert print photographs to digital format).

Storage for digital objects is also becoming increasingly accessible – hard drives (external and internal) are relatively cheap and easily configured. In this document there are general guidelines for users on how to take advantage of these storage options for backup purposes.

With the increasing take-up of user friendly digital technologies, the number and size of digital objects is growing exponentially. So too is the opportunity for the Library to collect and preserve these objects.

A guide to describing your objects (creating metadata) is included in this document with the aim to provide an introduction to the importance of describing objects. Your descriptions will help you locate your digital objects and provide information about their creation and content for future generations. This section aims to provide some helpful hints for you to describe your collection.

General Rules and Guides

This document has been created to provide a guide for content creators. The standards for digital capture and/or reformatting throughout this document are presented as a **minimum** standard, and where possible, higher resolutions/bit rates/pixels per inch/etc should be employed.

The Library strongly advises you **do not** reformat your original digital objects into the standards below. For example, you should not directly convert a 640x480 pixels JPEG image into a TIFF file. If you want to make changes to an object, do it to a copy, and keep the master in its original format.

The Library is investigating the use of XENA, a freely available, Open Source piece of software produced by the National Archives of Australia, to convert files to an OpenDocument format.

This document is subject to periodic review. As technologies change, the Library believes it should remain flexible to adapt to standards.

Always adhere to Copyright restrictions when digitising objects

Some General Hints for Preserving Digital Objects

Digital hardware, software, file formats and media are constantly being updated by manufacturers. This eventually causes superseded systems to become obsolete. Examples are the word-processing software Wordstar and also 5.25 floppy disks. People who wish to preserve their digital files for themselves and future generations should be vigilant in:

- retaining the manufacturers' software needed to read original digital files
- maintaining systematic metadata
- ensuring that preservation copies are made of original digital files in formats which are recognizable by open source software (i.e. are not brand-specific)
- running a regular backup regime so that copies exist of the preservation archive of digital files
- keeping several copies of the preservation archive in separate locations as a guard against loss through disaster
- periodically checking that their file formats such as tiff and PDF/A and storage media are still current
- copying their preservation files to new formats when current formats are approaching obsolescence
- copying their preservation archive to new storage media when their current storage media such as CDs are approaching obsolescence
- periodically checking that they can still access the information in the digital files
- checking that they will still be able to access their digital files when planning an upgrade to new computer equipment and/or software.

Standards for Digital Capture and Reformatting

Text

Creating a digital text

For born digital objects, the Library will collect the master document in its unaltered format, and aim to create a primary preservation object in PDF/A (Portable Document Format for Archiving) and an additional object in OpenDocument format.

To help ensure long term access, it is strongly suggested when you create a text document that you save a copy in a preservation format. This will allow you to minimise software migration issues. For example, if you are using Microsoft Word 2000, use a PDF converter or other format converter to create an archival copy. This archival copy should not rely on proprietary software to open it. Even when you have this copy, you should keep the original object in its original file format.

PDF/A is the current suggested file format for preservation. PDF/A is designed for the long-term preservation of text documents and is the first part of the standard documented in ISO 19005-1:2005(E). PDF/A is a subset of PDF, leaving out PDF features not suited to long-term archiving.

The OpenDocument format is an alternate option. It is an international standard (ISO/IEC26300:2006 Open Document Format for Office Applications v1.0.) that was developed for long-term preservation of electronic office documents.

Did you know?

- Open Office (v2.4+), a free open source office suite, allows users to save files in both suggested archival formats. Details about Open Office and where to download it are included in **Appendix Two**
- Microsoft Word 2007 also has an Add-in that allows documents to be saved in PDF/A format. Information on downloading this Add-in can be found in **Appendix Two**

Creating searchable text

If you intend to scan a text document (for example, if you wanted to digitise your Grandmother's cookbook), use the scanning standards specified for 'Images' below. You may also want to investigate Optical Character Recognition (OCR). OCR is the mechanical or electronic translation of images of handwritten, typewritten or printed text (usually captured by a scanner) into machine editable text. Further information is included in **Appendix Two**.

OCR is most effective on printed (typed) materials. Manual transcription of material is an additional option for hard to read/non-OCR friendly text. By providing OCR-ed text, you make the object searchable and more easily readable.

Did you know?

- The Microsoft Office Document Scanning software that regularly comes bundled with a Microsoft Suite has an OCR option (see **Appendix Two**)

Images

Taking a digital photograph

We believe that a digital camera with a resolution of at least 6 megapixels is best for capturing digital pictures. Many compact digital cameras capture images in JPEG format only. These JPEG images should be retained, unedited, as the master copy. For cameras with a choice of formats, the following settings are recommended (in order of preference); RAW, TIFF, JPEG.

If possible, set your digital camera to capture the image in the native or raw format (for example Canon uses the file extension .crw or .cr2, Olympus uses .orf, Kodak uses .dcr or .drf, and so on). If you are not able to set this on your camera, the next best option is to capture the image in an uncompressed TIFF format. If the raw format or TIFF is not a setting available on your camera, the JPEG standard is recommended.

If possible, select an uncompressed format using the highest resolution setting. This may reduce the number of images you can take on your camera but will allow these images to be of the highest quality. You may have to refer to the User Manual which came with your camera on how, or if it is possible, to select these settings for your camera.

Note:

- If you use the native or camera raw format for your camera original images, ensure you do not misplace the conversion software which comes with the camera. You will

need it to convert the RAW files into viewable TIFF or JPEG images. Native or camera raw formats are usually brand specific so it is advisable to create a preservation set of your raw files in an open format such as uncompressed TIFF format to ensure that you have a viewable set of images should the conversion software become obsolete.

It is important to remember that whatever file format you use to take pictures, you should not convert the 'master' file to fit the above recommendations – i.e. if your camera only takes pictures in the JPEG format, retain the JPEG as the master even if you make a duplicate TIFF file. There is information held in the original file (technical metadata), which may be lost in the conversion process. Of course, you may convert **a copy** of the image for re-use (e.g. if you wish to duplicate a TIFF file as a compressed JPEG format to email it to a friend). Refer to the section below on Describing Your Objects for hints on how to help you identify if an object is a copy or a master.

To summarise, we recommend:

- your camera resolution is at least 6 megapixels
- your file format should be 'native format' or 'Camera RAW' files
- if 'native format' or RAW is unavailable, use TIFF
- if JPEG is the only option, you should select the maximum quality setting
- you should take your pictures at the Maximum quality (finest or best quality setting, with no compression)
- you should retain your conversion software (the software that came with your camera) if making "native format" or Camera RAW files and make a backup set of images in TIFF format
- you retain your original camera images as masters and make duplicates for editing and conversion.

Digital conversion (scanning)

If you intend to scan photographic negatives, prints or transparencies to create digital objects, we suggest you use the following settings on your scanner. The scan should be:

- An uncompressed TIFF file
- Transparency & negative, 35mm, with an **output resolution** of 3000 ppi (pixels per inch)
- Transparency, negative and prints larger than 35mm, with an **output resolution** of 2000 ppi (pixels per inch)
- 48-bit RGB for colour photographic images
- 16-bit Greyscale for black and white photographic images

The above output resolutions relate to the scanner's **optical resolution output.**
If your scanner cannot meet these output resolutions, go to the maximum optical resolution of the scanner, **do not use interpolated resolution**.

Note:

- You may have to refer to your scanner User manual to configure the settings above

Image manipulation

Where possible, avoid changing the master image. Once you have a copy of the master image (as is, high resolution), you can then create a copy of this for manipulation. The master image should include all original imperfections such as scratches, red-eye, creases, dust spots etc. This will help keep the authentic feel of the picture. Any 'tidying up' of the image should be done on a copy. Refer to the section below on Describing Your Objects for hints on how to help you identify if an object is a copy or a master.

Analogue to digital conversion by State Library
In general, the State Library will produce a set of five digital objects when capturing images from paper and photographic media. This includes published and manuscript material, transparencies, negatives and photographic prints.

The five digital objects are:
Preservation Master – uncompressed TIFF
Access Copy - uncompressed TIFF
Thumbnail Image – PNG file embedded for display in the online catalogue. (Thumbnails maximum of 150 pixels on the longest side)
Viewing Image – JPEG file for a larger view on the web (760 pixels longest side)
XENA – xml file (National Library of Australia preservation format)

Audio

Making a digital audio recording
When using a digital recorder, we recommend you make the recording using an uncompressed Broadcast Wave Format (BWF.wav) file in 48 kHz, 24 bit. The BWF format is like the standard digital audio WAV file but carries extra metadata. If your recording device cannot capture at this quality, the following lists the files and formats we suggest, in order of preference:

- WAV, AIFF or FLAC file @ 48 kHz, 24 bit
- MP3 - if this is the only option. The master file should be the unedited "born digital" file from the recorder. It should be 128kb/sec (or higher).

Digital Conversion
As per the previous section, if you wish to convert your analogue audio tapes to a digital format, we recommend you record the object as a Broadcast Wave Format (BWF .WAV) file in 48 kHz, 24 bit.

If it is not possible to create BWF files, an uncompressed WAV, AIFF or FLAC file @ 48 kHz, 24bit is recommended.

Analogue to digital conversion by State Library
The State Library converts analogue to digital recordings using BWF .wav for the master copy and MPEG 1-Layer 3 (MP3) for the access copy. As with all digital objects, the master will not be enhanced through editing (e.g. to remove clicks, hum and buzz noise).

The access copy presented on the Web may be edited to remove clicks, hum and buzz etc. This web copy will be an MP3.

The Library is currently investigating other derivative formats to suit various client needs. For now, we believe that the MP3 file format is easily downloadable via the internet, and has an acceptable sound quality.

Did you know?

- An example of Open Source audio capture/conversion software is available at http://audacity.sourceforge.net/

Moving Images

Making a digital video recording

Video footage can be created in any one of a complex and evolving range of different formats. Since introduction in the 1950s there have been around 70 different tape formats developed and utilised. There are also at least that many variations in digital video 'flavours'. This wide range of available formats makes the selection of a long-term preservation format very difficult. Each format requires evaluation for the best way to preserve the images and make them accessible. For best results, choose the highest quality setting on the video camera. If there is a 'native format' or 'camera RAW' format, select that.

It is vital that when you record or copy your vision to hard drive you consider what it may take to play back to future generations. To maintain ultimate quality then the camera's RAW or native format is going to be best. However, it is often not possible to share this with friends without specialist software or install CDs that may have come with the camera. It is highly recommended you look at widespread and fairly standardised formats such as DV, MPEG2, MPEG4, or old fashioned standards such as photojpeg quicktime or 'animation' quicktime along with storing your native footage.

Digital conversion

When converting analogue moving picture recordings we recommend your output master copy is in AVI format and is uncompressed.

Analogue to digital conversion by State Library

When converting analogue moving picture recordings to digital format, the State Library creates a Master copy in AVI (uncompressed) format and Access copies in MPEG2 DVD format.

Did you know?

- An example of Open Source video capture/conversion software is available at http://www.virtualdub.org/

Describing Your Objects (creating Metadata)

Metadata helps you describe and access an object, and is essentially data *about* data (objects). Metadata can include the technical information stored on your digital camera (such as the date and time an image was taken); distinct keywords and conventions used in the naming of digital files; or

more detailed descriptions about the context and content of the digital object. Metadata provides a way of indexing, accessing, preserving, and discovering digital resources.

Naming Convention – You should use a standard way of naming your objects. Though digital files include a date stamp, you may want to consider adding this to the name of the object, as the date stamp on a digital object is the date that the object is created or modified, which introduces issues if you've scanned an analogue photograph into a digital format.

Some examples are:
20081023_Mandurah_Holiday
[date of creation YYYYMMDD]_[location]_[subject]

By putting the date first, you can easily sort objects by date.

You may also find it helpful to add detail to your file names:
20081023_MandurahHoliday_Kids
20081023_MandurahHoliday_ViewFromHotel
20081023_MandurahHoliday_MyFirstFish
20081023_MandurahHoliday_GrandmaGrandadJoe

[date of creation YYYYMMDD]_[location]_[subject]_[detail]

The effort you spend at this stage of the process will make it much easier for you to accurately find a specific image from your collection.

Consider the above example as opposed to:
0001.jpeg
0002.jpeg
0003.jpeg
0004.jpeg

Which one was that picture of me catching my first fish?

More detailed descriptions
You may wish to include more detailed metadata about your images. Some examples of more detailed information you may want to provide are:

Version control - Indicators in object names to identify if the object is the master version or a copy (indicated by an 'M' or 'C' in the names below):

20081023_**M**_MandurahHoliday_GrandmaGrandadJoe.tiff
*Note the '**M**', indicating that this is the Master Copy*

20081023_**C**_MandurahHoliday_GrandmaGrandadJoe.tiff
*Note the '**C**', indicating that this is a Copy of the Master, and can be edited*

Location identification – Used for identifying the Location where the object was created and for

Providing Contextual information (notes) e.g. This was the first holiday the kids had with Grandma and Grandpa/ Sam's first swim in the ocean, etc

Some of this metadata may be held in the folder name – e.g.
20081208_Birthday_Aunty_Wilma_North_St

Even more detailed descriptions
In addition to distinct and consistent file names as outlined above, you may also wish to develop more detailed descriptions of your digital objects. More detailed metadata can help you to search for and identify objects, remember the significance of the event recorded in a digital image, and record information about the conditions of use of the digital object. This metadata, though ideally stored as part of the object, may be held in a separate file to the object – for example, in a text document saved in the same directory as a group of pictures.

Did you know?

- An example of Open Source software that can read and edit metadata embedded in files can be found here: http://www.sno.phy.queensu.ca/~phil/exiftool/

There are many metadata schemas available and they range from simple to complex formats. The choice of a schema depends on how and who will be using the schema. The Dublin Core metadata schema is a simple standard that is easily applicable to personal collections and allows individuals to produce their own metadata with no extensive training.

Your metadata may include some or all of the following summarised Dublin Core elements:

Title	The name or title given to your digital object
Rights	Information about rights held in and over the digital object. If you own the rights to the digital object, you may wish to apply a Creative Commons license to your digital object and record the conditions of that license in this field.
Subject	The subject of your digital object. Use keywords, or a controlled vocabulary to represent the subject.
Description	An explanation of your digital object. The description may include an abstract or other summary.
Creator	The person or organisation that created the digital object (e.g. a digital copy of a photograph your grandmother took in 1926)
Contributor	A body responsible for having some input to the object. Examples of a Contributor include a person, an organization, or a service. (e.g. your grandmother who took the original analogue image in 1926 that you've now digitised).

Date	A date or period of time associated with the life-cycle of the digital object. This may include the date the image was taken; the date the image was copied and image corrected; or the range of dates the oral history was taken (e.g. an oral history recording with your great aunt that was made over 3 days).
Type	The nature or genre of your digital object (e.g. Collection, Image, Moving Image, Sound, Still image, Text).
Format	The file format, physical medium, or size and duration dimensions of the resource.
Source	If your object is not 'born digital', the source from which your digital object is derived. (e.g. the original hard copy cookbook scanned and turned into a digital object).
Coverage	The space and/or time subject of the object. The space may be the geographic location (e.g. a place name, or GPS coordinates); the time may be a date or range of dates covered by the object (e.g. an oral history interview about your father's experiences during the Vietnam war)
Relation	A related resource to your digital object. (e.g. the hard copy transcript of the digitally recorded interview)
Language	The language of the resource (e.g. an interview conducted in Polish and English; an image of my grandmother's cookbook with annotations in Greek).
Publisher	The person or organisation who published the object (if it is published)
Identifier	A unique reference to the resource within a given context. This may be the name you give to your file

There are not many limitations to including detailed metadata; the most common of which is time required and following a standard method of describing your collections. The more information you can provide, the easier it will be for yourself and others to find specific objects and has the added benefit of making it easier to recall details about events.

Data security

In brief, where you can, back up your data and keep a copy separately to your main storage.

For example, if you keep your digital pictures on your home computer, make a copy of them and keep that copy elsewhere (a family member's house, in your office at work, etc). By keeping a separate copy, if there is a disaster at your primary storage site (e.g. your house floods, and the

home PC with all your digital recordings of your music is damaged), you have a copy that you can recover your data from.

You may also wish to consider backing up your data on a different media – e.g. keeping a printed copy of your pictures. If you decide to use a different softcopy media (e.g. burn your pictures on to CD), make sure you periodically check the disc to ensure the files are still accessible.

There are many options to simplify this process, for example 'one touch' or automated backup devices or software which can be configured to back up a directory or incrementally add backups as required, or to CD, DVD or external hard drive.

Regularly backup your collection.
There are a couple of methods to do this:
-Make a complete backup of your data i.e. file #1 through to file 1000 every time. Depending on the media that you backup to, this can be an expensive process.

-Add to your backups incrementally
January – file #1>file 500
Feb- file 501>file 600 (files created in January)
March – file 601>700(files created in Feb)

A combination of the two:
Every January, take a complete backup of your files
Monthly after that, increment your backups.

The question to ask yourself is "When the hard drive that I am keeping my data on fails (and it will fail – it's just a question of when), how much data can I afford to lose?"

A real life example:
Sarah regularly records movie files of her 3year old child. She stores these files on the hard drive of her home PC. She has two external hard drives for backing up data to (she has named them HD-A and HD-B). Sarah has decided that she should back up her files every month.

Every month (on the 1st of the month), she backs up all her files to HD-A. She then takes that hard drive to her mother's house. She brings home HD-B (which has been at her mum's since last month) and repeats the backup process onto this other drive. When she's done, she has three complete copies of her movie files – one on her PC, one on HD-B, and one on HD-A (now at her mum's).

Remember, this process takes commitment. Keeping a regular backup schedule will make this easier (and you're less likely to forget). The cost of not doing this could be that you lose everything!

Appendix 1:
Glossary of terms

AIFF	Audio Interchange File Format (AIFF) is an audio file format standard used for storing sound data for personal computers and other electronic audio devices. The format was co-developed by Apple Computer in 1988 [1] based on Electronic Arts' Interchange File Format (IFF, widely used on Amiga systems) and is most commonly used on Apple Macintosh computer systems. Audio Interchange File Format, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Audio_Interchange_File_Format
Analogue	A non-digital item. Examples include printed materials, photographic negatives, audio cassette tapes, etc.
Audio data compression	Audio compression is a form of data compression designed to reduce the size of audio files. Audio compression (data), in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Audio_data_compression
AVI	Audio Video Interleave is a multimedia container format introduced by Microsoft in November 1992 as part of its Video for Windows technology. AVI files can contain both audio and video data in a file container that allows synchronous audio-with-video playback. AVI, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/AVI
Bit	A bit is a binary digit, the smallest increment of data on a computer. A bit can hold only one of two values: 0 or 1, corresponding to the electrical values of off or on, respectively. Because bits are so small, you rarely work with information one bit at a time. Bits are usually assembled into a group of eight to form a byte. A byte contains enough information to store a single ASCII character, like "h". Retrieved 10 February 2009 from http://kb.iu.edu/data/ackw.html
Born Digital	Material that has originated in a digital form (digital camera, digital video, digital recorder, etc), should be submitted in the original format produced by the device. Details of the brand and model of device (camera type and any special software required to view the images, etc) should also be documented or submitted with the material.

BWF	Broadcast Wave Format (BWF) is an extension of the popular Microsoft /IBM WAVE audio format and is the recording format of most file-based non-linear digital recorders used for motion picture and television production. BWF, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/BWF
Creative Commons Licensing	A free, structured way creators can assign freedoms to their work http://creativecommons.org/
Data compression	Data compression, the process of encoding information using fewer bits See also Image compression, Audio data compression, and Video compression
Derivative	A copy of your digital object – may be in a different format (e.g. a print copy of your digital photograph).
Digital object	An object in a digital form (i.e. not a print)
File format	The type of file – examples include .jpeg, .avi, .doc, etc.
FLAC	Free Lossless Audio Codec (FLAC) is a file format for lossless audio data compression. Being lossless, FLAC does not remove information from the audio stream, as lossy compression formats such as MP3, AAC, and Vorbis do. Free Lossless Audio Codec, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Free_Lossless_Audio_Codec
Image compression	Image compression is the application of data compression on digital images. In effect, the objective is to reduce redundancy of the image data in order to be able to store or transmit data in an efficient form. Image compression, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Image_compression
Interpolated resolution	An enhanced resolution of a scanning device that is computed using a software algorithm. Also called the "digital resolution," it makes an image appear as if it were scanned at a higher resolution. An interpolated resolution is considerably greater than the optical resolution, which is the inherent physical resolution of the device. Depending on the contents of the image and the scanning algorithm, an interpolated, or enhanced, resolution can improve or degrade the original. Interpolated resolution technology, in Answers.com, Retrieved 12 December 2008 from http://www.answers.com/topic/interpolated-resolution-technology

JPEG	In computing, JPEG a commonly used method of compression for photographic images. The degree of compression can be adjusted, allowing a selectable trade-off between storage size and image quality. JPEG typically achieves 10 to 1 compression with little perceptible loss in image quality. JPEG, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/JPEG
Master	The original object, unedited, in its original format.
Metadata	Metadata (meta data, or sometimes metainformation) is "data about data", of any sort in any media. An item of metadata may describe an individual datum, or content item, or a collection of data including multiple content items and hierarchical levels, for example a database schema. In data processing, metadata is definitional data that provides information about or documentation of other data managed within an application or environment. The term should be used with caution as all data is about something, and is therefore metadata. Metadata, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Metadata
MP3	MPEG-1 Audio Layer 3, more commonly referred to as MP3, is a digital audio encoding format using a form of lossy data compression. It is a common audio format for consumer audio storage, as well as a de facto standard encoding for the transfer and playback of music on digital audio players. MP3, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/MP3
MPEG 1 - Layer 3 (see MP3)	MP3, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/MP3
MPEG2	MPEG-2 is a standard for "the generic coding of moving pictures and associated audio information".[1] It describes a combination of lossy video compression and lossy audio compression (audio data compression) methods which permit storage and transmission of movies using currently available storage media and transmission bandwidth. MPEG-2, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/MPEG-2
Native format	The format which is created by the recording device – generally these are proprietary (which means you need special software to view the files in this format). Native formats often hold data which is not accessible in other formats.

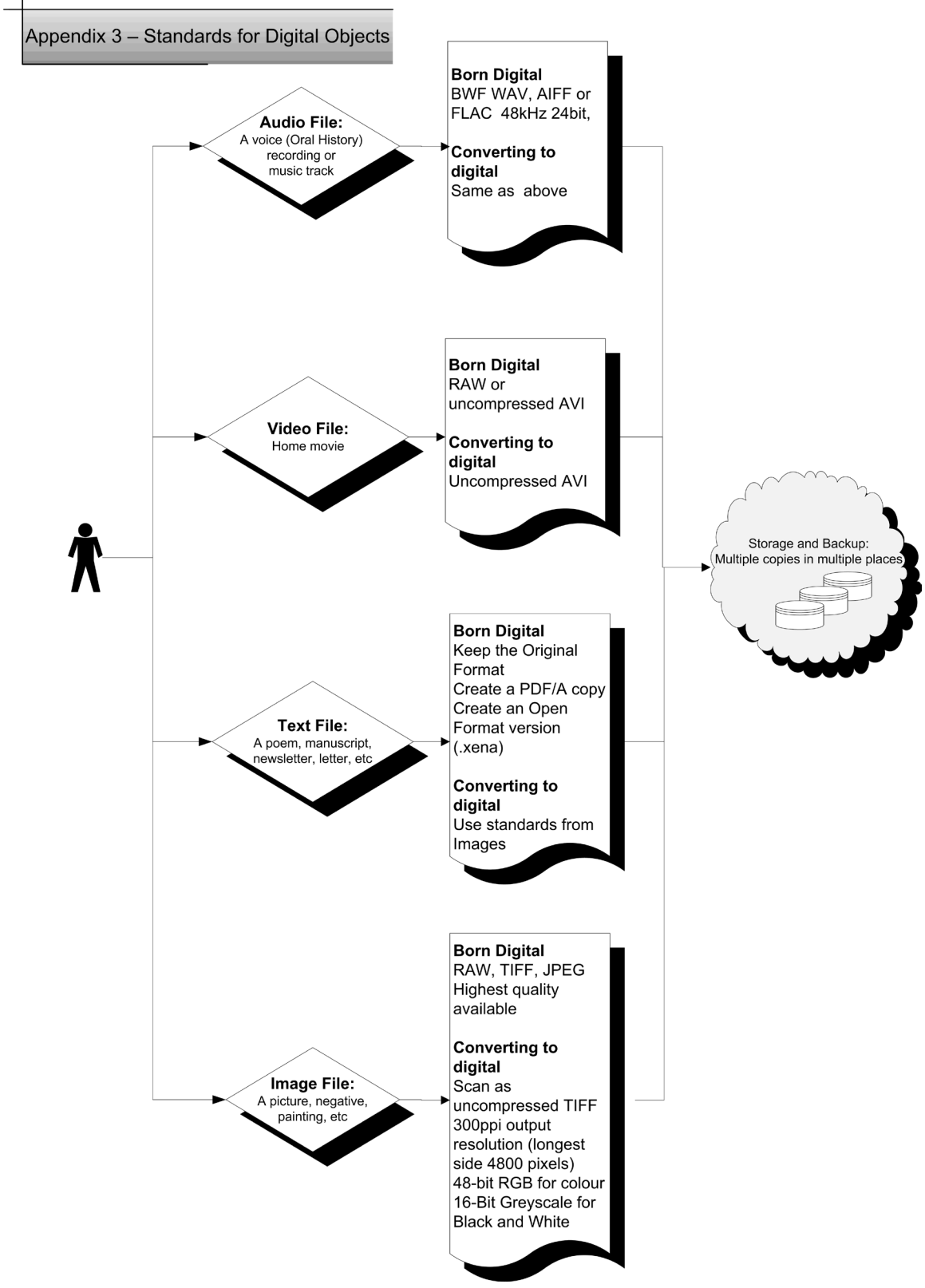
OCR	Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic translation of images of handwritten, typewritten or printed text (usually captured by a scanner) into machine-editable text. Optical Character Recognition 20/11/08, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Optical_character_recognition 20/11/08 For an example of Open Source OCR software, see: http://code.google.com/p/tesseract-ocr/, Retrieved 16 December 2008
Open format	An open format is a published specification for storing digital data, usually maintained by a standards organization, which basically can be used and implemented by anyone. For example, an open format can be 'implementable' by both proprietary and free and open source software, using the typical licences used by each. In contrast to open formats, proprietary formats are controlled and defined by private interests. Open format, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Open_format
Optical resolution	Optical resolution describes the ability of an imaging system to resolve detail in the object that is being imaged. Optical resolution, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Optical_resolution
Open Source software	Generally free to use software that allows users to improve and change the source code.
Output resolution	Output resolution refers to the number of image pixels in each linear inch of the picture. Printing – resolving resolution, reprinted from PCPhoto Magazine, April 2001 in University of Wisconsin, Platteville, Retrieved 12 December 2008 from http://www.uwplatt.edu/~covert/Newsletter/printing/ResolvingResolution.htm
PDF	Portable Document Format (PDF) is a file format created by Adobe Systems in 1993 for document exchange. PDF is used for representing two-dimensional documents in a manner independent of the application software, hardware, and operating system Portable Document Format – 20/11/08, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Portable_Document_Format_-_20/11/08

PDF-A	This standard defines a format (PDF/A) for the long-term archiving of electronic documents and is based on the PDF Reference Version 1.4 from Adobe Systems Inc. (implemented in Adobe Acrobat 5). PDF/A is in fact a subset of PDF, leaving out PDF features not suited to long-term archiving. PDF/A 20/11/08, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/PDF/A 20/11/08
Pixel	In digital imaging, a pixel (picture element) is the smallest piece of information in an image. Pixels are normally arranged in a regular 2-dimensional grid, and are often represented using dots, squares, or rectangles. Each pixel is a sample of an original image, where more samples typically provide a more accurate representation of the original. Pixel, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Pixel
PNG	Portable Network Graphics (PNG) is a bitmapped image format that employs lossless data compression. PNG was created to improve upon and replace GIF (Graphics Interchange Format) as an image-file format not requiring a patent license. http://en.wikipedia.org/wiki/Portable_Network_Graphics
Raw	A raw image file contains minimally processed data from the image sensor of a digital camera or image scanner. Raw files are so named because they are not yet processed and ready to be used with a bitmap graphics editor or printed. http://en.wikipedia.org/wiki/RAW_image_format 20/11/08
TIFF	Tagged Image File Format (abbreviated TIFF) is a file format for storing images, including photographs and line art. It is now under the control of Adobe Systems. Tagged Image File Format (TIFF) is a variable-resolution bitmapped image format developed by Aldus (now part of Adobe) in 1986. TIFF is very common for transporting colour or grey-scale images into page layout applications, but is less suited to delivering web content. TIFF, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/TIFF
Video compression	Video compression refers to reducing the quantity of data used to represent digital video images, and is a straightforward combination of image compression and motion compensation. Video compression, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/Video_compression

WAV	WAV (or WAVE), short for Waveform audio format, is a Microsoft and IBM audio file format standard for storing an audio bitstream on PCs. WAV, in Wikipedia, the free encyclopedia, Retrieved 12 December 2008 from http://en.wikipedia.org/wiki/WAV

Appendix 2:
Links to other sites with good digitisation information
Storage and Preservation
National Archives of Australia (http://www.naa.gov.au).
Xena
http://xena.sourceforge.net/
Microsoft Add-in to convert to PDF
http://www.microsoft.com/downloads/details.aspx?FamilyId=F1FC413C-6D89-4F15-991B-63B07BA5F2E5&displaylang=en
Open Office – Details and Download location
http://download.openoffice.org/
OCR
Microsoft Office Document Scanning
http://office.microsoft.com/en-us/help/HP010771031033.aspx
Start>Programs>Microsoft Office>Office Tools> Microsoft Office Document Scanning
Optical Character Recognition
http://en.wikipedia.org/wiki/Optical_character_recognition
Creating RAW images
http://www.tasi.ac.uk/advice/creating/raw-format.html
Uncompressed TIFF
http://www.dpreview.com/learn/?/key=TIFF
Output resolution
http://www.bowhaus.com/news/inputoutput.php4
Resolution
http://graphicssoft.about.com/od/glossary/g/resolution.htm
16 bit greyscale
http://www.imaging-components.com/Imaging-Components/Image-Processing/bits-per-pixel.shtm
Introduction to Metadata
http://www.getty.edu/research/conducting_research/standards/intrometadata/
Dublin Core (metadata schema)
Dublin Core Metadata Element Set, Version 1.1: http://dublincore.org/documents/dces/
Toolkit to select a metadata schema
http://www.rubric.edu.au/packages/RUBRIC_Toolkit/docs/Metadata.htm
Rights
Creative Commons: http://creativecommons.org/

Appendix 3


Appendix 3 – Standards for Digital Objects
Audio File:
A voice (Oral History) recording or music track
Born Digital
BWF WAV, AIFF or FLAC 48kHz 24bit,
Converting to digital
Same as above
Video File:
Home movie
Born Digital
RAW or uncompressed AVI
Converting to digital
Uncompressed AVI
Storage and Backup:
Multiple copies in multiple places
Text File:
A poem, manuscript, newsletter, letter, etc
Born Digital
Keep the Original Format
Create a PDF/A copy
Create an Open Format version (.xena)
Converting to digital
Use standards from Images
Image File:
A picture, negative, painting, etc
Born Digital
RAW, TIFF, JPEG
Highest quality available
Converting to digital
Scan as uncompressed TIFF
300ppi output resolution (longest side 4800 pixels)
48-bit RGB for colour
16-Bit Greyscale for Black and White